Improvements to the Predictive Model

Jonathan Cohen, ICF Consulting

Gary Whitten, Smog Reyes

2 August, 2005

OBJECTIVE

- To best represent the impact of gasoline fuel variations on the total on-road fleet emissions using the available database

Historical Overview
NOx Tech 5 (1994 and later)

- Limited database in 2001
- CaRFG3 excludes Alliance, AIAM, Honda oxygen fuel emissions data
- CaRFG3 assumes same NOx response as Tech 4
- SWRI/EPA assumed no NOx increase

Historical Overview
NOx Tech 4 (1985 to 1993)

- CaRFG2 built on stepwise approach
  - Main effects: AR, OL, OX, RV, SU, T5, T9
  - Interactions: T9SU, OXOX, AROX, RVOX, RVRV
- CaRFG3 initially used stepwise approach
  - Unacceptable to stakeholders
  - Too many terms
  - Unexpected responses

Historical Overview
NOx Tech 4, continued

- CaRFG3 fits same terms as CARFG2
  - Signif. Interactions: T9SU, OXOX, AROX
- SWRI/EPA reconstruct model (2000)
  - Include “high influence” vehicles
  - Renormalize fuel parameters to mean 0, std. dev. 1
  - Stepwise approach starting with main fuel effects
  - Candidates include: fuel × fuel interactions,
  - “High-emitter” main effect,
    - HC > 2 × Standard (0.82 g/mi) or
    - CO > 2 × Standard (6.8 g/mi)
  - High-emitter × Main fuel effect interactions

Dual and Weighted Models

- Graboski, Cohen and Pollack, 2000
- Whitten and Cohen, 2005
- Dual model fits separate mixed models to:
  - Normal emitters (EMFAC NOx normal)
  - “Higher” emitters (EMFAC NOx moderate, high, very high, super)
- Weighted model
  - Vehicle weight $\propto$ NOx emissions (2005) for EMFAC category (normal, moderate, high, or very high; no super-emitters)
  - Approximate analysis assuming same test fuels on each vehicle
- Dual model fits statistically significantly better
- Weighted and Version 1 dual models give similar oxygen-NOx predictions

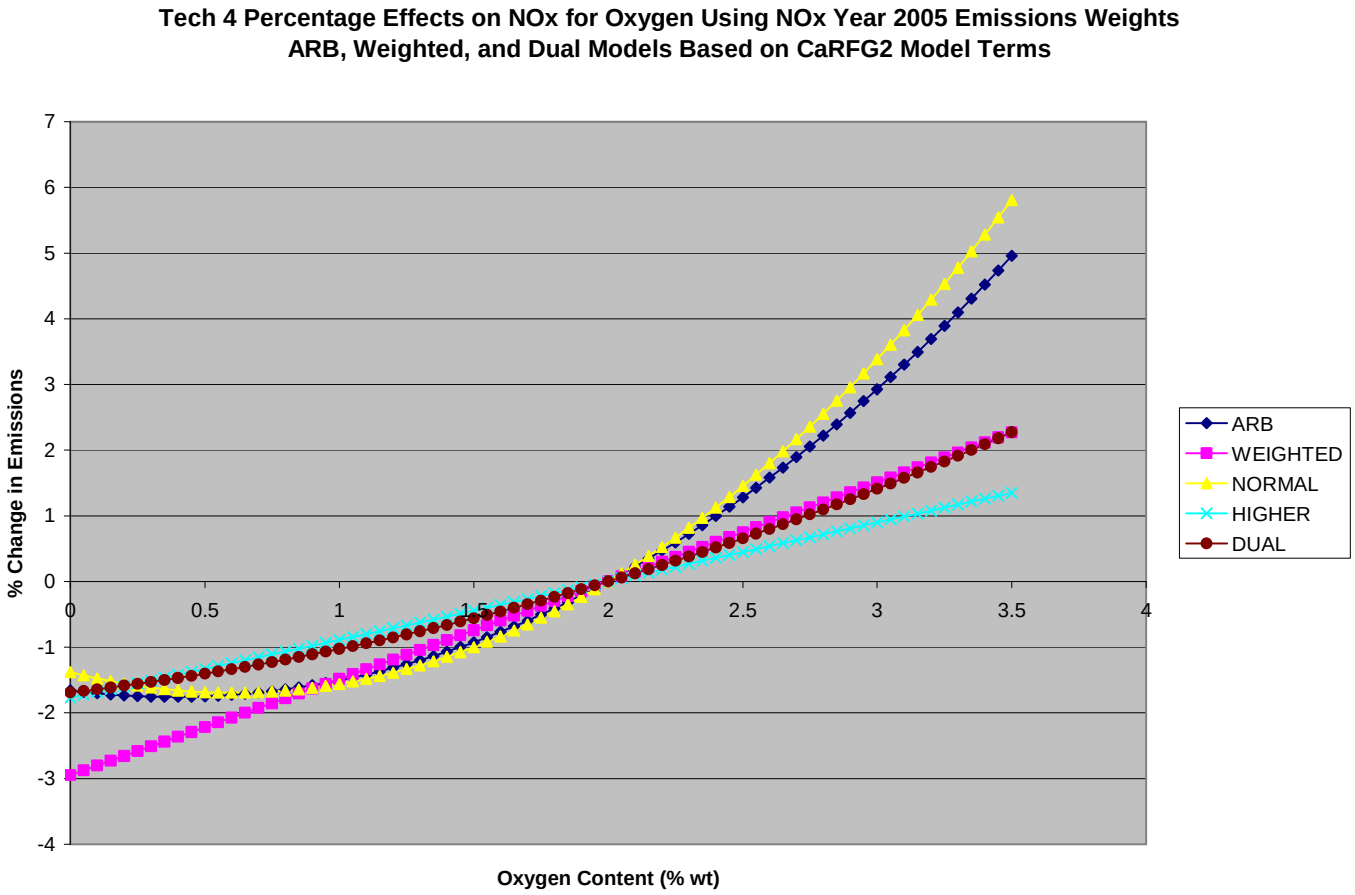
ICF Dual Model Features

- Higher emitters based on NOx, not THC/CO
- Higher emitters defined as > 1 × std,
  not > 2 × std
- Separate models, not higher emitter interaction terms
- ML, not REML
- Include fuel × fuel interactions as candidate terms for higher emitters
- Three versions

Dual Model Versions

- Version 1: Limited to same candidate interaction terms as CaRFG2 and CaRFG3
- Version 2: Stepwise approach with p-values:
  - Use all fuel × fuel interactions as candidate terms
  - Add terms based on most significant p-values < 0.05
  - Subtract non-significant terms
- Version 3: Stepwise approach with AICC:
  - Use all fuel × fuel interactions as candidate terms
  - Akaike Information Criteria Corrected (AICC):
    - Log-likelihood goodness-of-fit measure penalized for # parameters (“corrected” for small sample sizes)
  - Add terms based on best AICC improvement

Version 1 and Weighted Model Results for Oxygen



Higher Emitters Based on NOx, Not THC/CO

- High NOx emitters ≠ High THC or CO emitters
- Often, High THC ⇒ Low NOx (Rich fuel mixture)
- Therefore, more logical to use NOx levels for modeling NOx emissions

Higher Emitters Based on 1×Std, Not 2×Std

- Goal is to produce best statistical model for the emissions data
- Five EMFAC categories
  - Normal (below std)
  - Moderate (1 to 2 × Std)
  - High (2 to 3 × Std)
  - Very High (3 to 4.5 × Std)
  - Super (> 4.5 × Std). None in test fleet
- Inadequate data to separately model High + Very High emitters
- Adequate data to separately model Moderate + High + Very High emitters = Higher emitters
- Higher emitters show different fuel responses than normal emitters

Separate Models, Not Higher Emitter Interaction Terms

- Assume 2 Intercept random effect terms:
  - Intercept (Int),
  - Intercept × Higher (Int × HI) interaction
- Implies V = Variance (veh mean) =
  - Var (Int) + Var (Int × HI), for higher emitters
  - Var (Int) for normal emitters
- Implies V (higher) ≥ V (normal)
- BUT ICF Dual model shows < not ≥

Separate Models, Not Higher Emitter Interaction Terms, Ctd

- Assume 2 OX random effect terms, OX and OX × HI
- W = Var (Veh OX effect) across vehicles
- Implies W (higher) ≥ W (normal)
- BUT ICF Dual model shows < not ≥
- Similarly for other main fuel effects

Separate Models, Not Higher Emitter Interaction Terms, Ctd

- Higher Emitter Interaction Terms approach has one error variance, σ^2
- Implies σ^2 (higher) = σ^2 (normal)
- ICF Dual model has two error variances
- BUT ICF Dual model shows σ^2 (higher) < σ^2 (normal)

ML, not REML

- ML = Maximum Likelihood
- REML = Restricted Maximum Likelihood
- Usually ML and REML give similar coefficients
- ML allows us compare models using:
  - Likelihood ratio test
    - 2 {Log-Lik (more terms) – Log-Lik (fewer terms)} $\cong \chi^2$
  - AIC(C) Goodness-of-fit
    - AICC = Log-Lik – f(no. of terms)
- Inapplicable for REML since models to be compared have different covariance structures

Three Dual Models

- Version 1: Same candidate interaction terms as CaRFG2 and CaRFG3
  - Fuel parameters renormalized using all data
  - Fuel Means ≅ 0, Std Devs ≅ 1
- Version 2: Stepwise approach with p-values
- Version 3: Stepwise approach with AICC
  - Fuel parameters separately renormalized for normal, higher
  - Fuel Means = 0, Std Devs = 1
- No impact of renormalization on fixed effects model
  - Same statistical model – re-parameterized
- Some impact on random effects model

Interaction Terms

- Version 1 Normal: T9SU, OXOX, AROX
- Version 1 Higher: None
- Version 1 Single: T9SU, OXOX, AROX
- Version 2 Normal: OXOX, SUOX, T5T5, ARSU, ARAR, OLOL, RVT9, T9AR, RVOL, T9OX
- Version 2 Higher: OLOX, SUSU, RVOX
- Version 2 Single: OXOX, SUOX, T5T5, ARSU, ARAR, OLOL, RVT9, T9AR, RVOL, RVSU
- Version 3 Normal:T9OX, OXOX, SUOX, T5T5, T9AR, RVAR, RVOX, OLOL, RVT9, T9SU, T5OL
- Version 3 Higher: RVOX, T9OL, RVT5, RVSU, RVT9, AROX, AROL, T5OL, RVRV
- Version 3 Single: SUOX, T5OX, T5T5, RVT9, RVOL, T5OL, ARAR, OXOX, T9AR, OLOL, RVSU, OLSU

Dual Versus Single Model Statistical Comparison

- Version 1 and 2: Use ML to compare
  - Log-Lik (Normal) + Log-Lik (Higher) vs.
  - Log-Lik (Single)
- For formal likelihood ratio test, single model needs to be a special case of dual model (normal + higher)
- If there are any interaction terms in the single model but not in the normal model, add them to the normal model. Ditto for the higher model.
- For Versions 2 and 3, use an overall renormalization rather than separate normal and higher fuel parameter renormalizations.

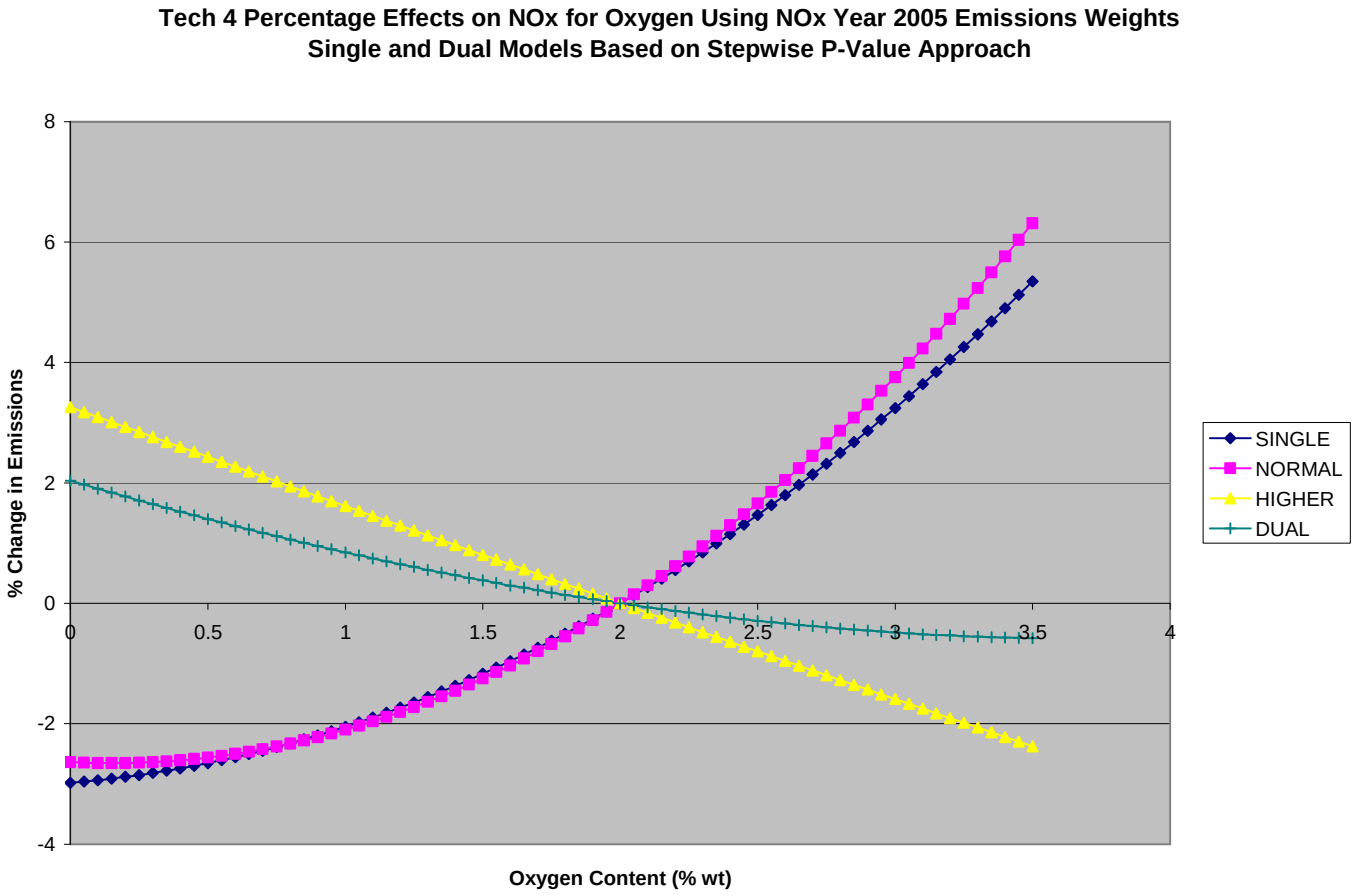
Dual Versus Single Model, Continued

- Single vs Dual:
  - Test different vehicle and/or fuel responses, i.e.
  - Are the average Log (NOx) levels different for higher emitters and/or
  - Are the fuel effects on NOx different for higher emitters?
- Single with higher emitter intercept vs. Dual:
  - Test different fuel responses (given different NOX levels)
- Version 3: Use ML to compare AICC

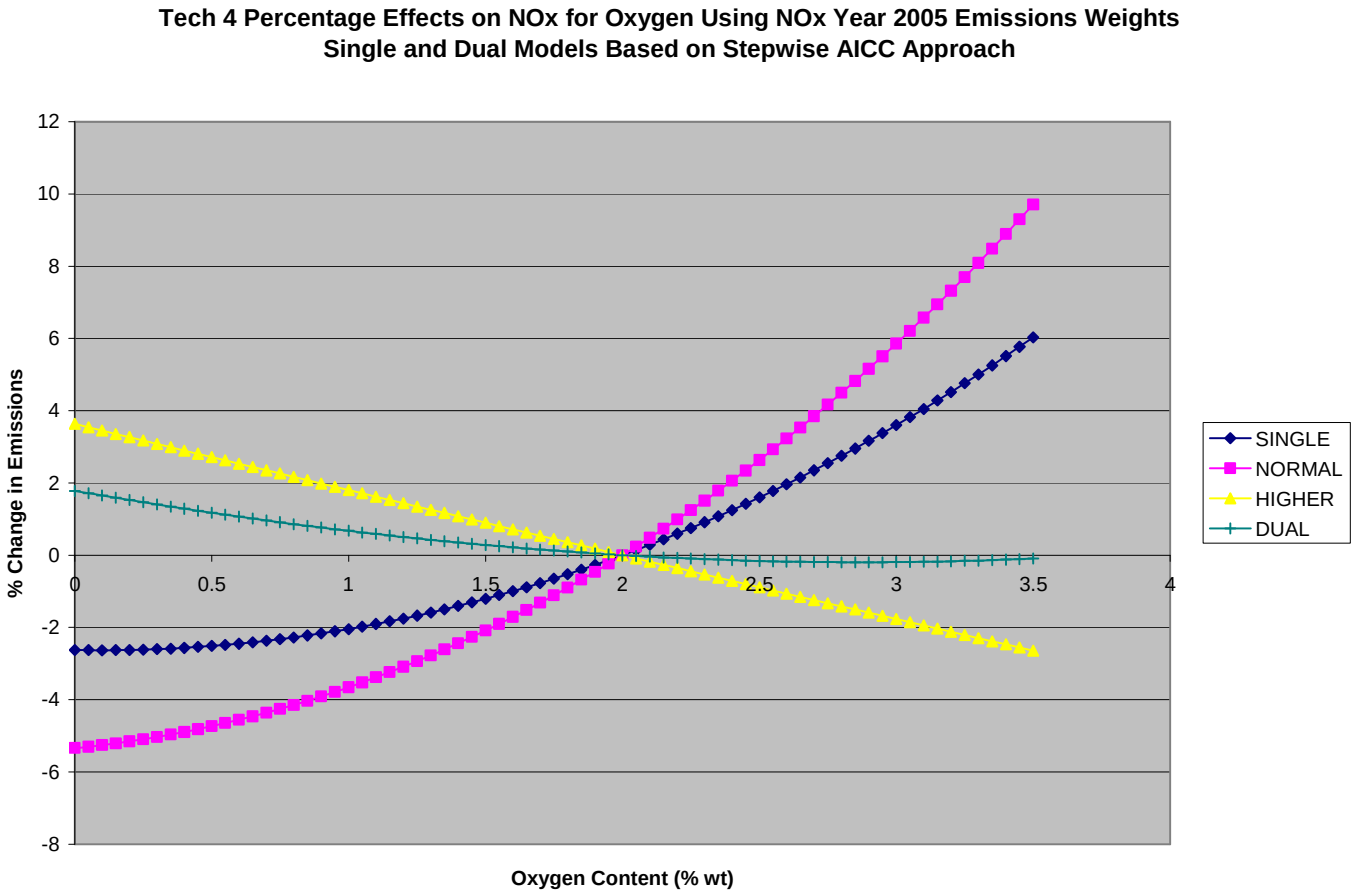
Dual Models Fit Better

Version	Test dual vs single model:	ΔAICC	P-value
1	Veh and Fuel effects	732.4	$< 10^{-10}$
	Fuel effects	266.4	$< 10^{-10}$
2	Veh and Fuel effects	752.1	$< 10^{-10}$
	Fuel effects	287.7	$< 10^{-10}$
3	Veh and Fuel effects	ΔAICC = 714.5	
	Fuel effects	ΔAICC = 253.8	

Version 2 Results for Oxygen



Version 3 Results for Oxygen



Appendix: Version 1 and Weighted Model Results for Oxygen on THC